Analyzing the Neighbourhoods in Mumbai for Starting a Restaurant

Applied Data Science Capstone Project

By: Tanveen Kaur

Date: August 15, 2021

Table of Contents

Introduction

Mumbai is the financial capital of India and is one of the most densely populated cities in the world. It lies on the west coast of India and attracts heavy tourism from all over the globe every year. It is one of the major hubs of the world and is extremely diverse with people from various ethnicities residing here. The multi-cultural nature of the city of Mumbai has brought along with it numerous cuisines from all over the world. The people of India generally love food and I personally love to try different cuisines and experience different flavours. Thus, the aim of this project is to study the neighbourhoods in Mumbai to determine possible locations for starting a restaurant. This project can be useful for business owners and entrepreneurs who are looking to invest and open a restaurant in Mumbai. The main objective of this project is to carefully analyse appropriate data and find recommendations for the stakeholders.

Data Collection

The following data is required for the project:

1) Neighbourhood data of Mumbai
2) Geographical coordinates of Mumbai and all neighbourhoods in Mumbai
3) Venue data for neighbourhoods in Mumbai

Neighbourhoods Data

The data of the neighbourhoods in Mumbai was scraped from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Mumbai. The data is read into a pandas data frame using the read_html() method. The main reason for doing so is that the Wikipedia page provides a comprehensive and detailed table of the data which can easily be scraped using the read_html() method of pandas. The top 10 rows of the dataframe are shown in following figure.

	Neighborhood	Location	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.129300	72.843400
1	Chakala, Andheri	Western Suburbs	19.111388	72.860833
2	D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373
3	Four Bungalows	Andheri,Western Suburbs	19.124714	72.827210
4	Lokhandwala	Andheri,Western Suburbs	19.130815	72.829270
5	Marol	Andheri,Western Suburbs	19.119219	72.882743
6	Sahar	Andheri,Western Suburbs	19.098889	72.867222
7	Seven Bungalows	Andheri,Western Suburbs	19.129052	72.817018
8	Versova	Andheri,Western Suburbs	19.120000	72.820000
9	Mira Road	Mira-Bhayandar,Western Suburbs	19.284167	72.871111

Figure: Top 10 rows of Mumbai neighbourhood's data scraped from Wikipedia.

Geographical Coordinates

The geographical coordinates for Mumbai have been obtained from the GeoPy library in python. This data is relevant for plotting the map of Mumbai using the Folium library in python. The code for getting the geographical coordinates of Mumbai is shown in the following figure.

```
address = 'Mumbai, IN'
geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Mumbai are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinates of Mumbai are 19.0759899, 72.8773928.
```

Figure: Obtaining geographical coordinates of Mumbai.

The geocoder library in python has been used to obtain latitude and longitude data for various neighbourhoods in Mumbai. The coordinates of all neighbourhoods in Mumbai are used to check the accuracy of coordinates given on Wikipedia and replace them in our data frame if the absolute difference is more than 0.001. These refined coordinates are then further used for plotting neighbourhoods using the Folium library in python. Figure 3 shows the coordinates of neighbourhoods in Mumbai obtained from Wikipedia as 'Latitude' and 'Longitude' and those obtained from geocoder as 'Latitude1' and 'Longitude1'. Furthermore, it also shows the absolute difference between the two latitude columns and the two longitude columns as 'Latdiff' and 'Longdiff', respectively. Once again only the top 10 rows are shown.

	Neighborhood	Location	Latitude	Longitude	Latitude1	Longitude1	Latdiff	Longdiff
0	Amboli	Western Suburbs	19.1293	72.8464	19.1291	72.8464	0.00024	0.00304
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	19.1084	72.8623	0.003028	0.001497
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	19.1251	72.8325	0.000965	0.001107
3	Four Bungalows	Western Suburbs	19.1263	72.8243	19.1263	72.8243	0.001606	0.00288
4	Lokhandwala	Western Suburbs	19.1432	72.8249	19.1432	72.8249	0.012345	0.0044
5	Marol	Western Suburbs	19.1192	72.8827	19.1191	72.8828	0.000169	6.7e-05
6	Sahar	Western Suburbs	19.1027	72.8626	19.1027	72.8626	0.00376476	0.00464166
7	Seven Bungalows	Western Suburbs	19.1315	72.817	19.1315	72.8165	0.00240802	0.000558001
8	Versova	Western Suburbs	19.1377	72.8135	19.1377	72.8135	0.01769	0.00652
9	Mira Road	Western Suburbs	19.2657	72.8711	19.2657	72.8707	0.0184624	0.000418149

Figure: Absolute difference between latitude and longitude values obtained from Wikipedia and Geocoder.

The figure below shows the top 10 rows of the final Mumbai neighbourhoods data frame after replacing the latitude and longitude values as mentioned before and dropping unnecessary columns.

	Neighborhood	Location	Latitude	Longitude
0	Amboli	Western Suburbs	19.1293	72.8464
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623
2	D.N. Nagar	Western Suburbs	19.1241	72.8325
3	Four Bungalows	Western Suburbs	19.1263	72.8243
4	Lokhandwala	Western Suburbs	19.1432	72.8249
5	Marol	Western Suburbs	19.1192	72.8827
6	Sahar	Western Suburbs	19.1027	72.8626
7	Seven Bungalows	Western Suburbs	19.1315	72.817
8	Versova	Western Suburbs	19.1377	72.8135
9	Mira Road	Western Suburbs	19.2657	72.8711

Figure: Final Mumbai neighbourhoods dataframe.

Venue Data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighbourhoods in Mumbai and is used to study the popular venues of different neighbourhoods as well as build the unsupervised learning model to cluster neighbourhoods. The venue recommendations of all neighbourhoods were obtained with a limit of 200, that is, maximum of 200 venue recommendations per neighbourhood and a radius of 1 km around the neighbourhood's geographical coordinates. Figure below shows the top 10 rows depicting the results obtained after cleaning the data from Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.1293	72.84644	Cafe Arfa	19.128930	72.847140	Indian Restaurant
1	Amboli	19.1293	72.84644	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
2	Amboli	19.1293	72.84644	Shawarma Factory	19.124591	72.840398	Falafel Restaurant
3	Amboli	19.1293	72.84644	Jaffer Bhai's Delhi Darbar	19.137714	72.845909	Mughlai Restaurant
4	Amboli	19.1293	72.84644	Narayan Sandwich	19.121398	72.850270	Sandwich Place
5	Amboli	19.1293	72.84644	Persia Darbar	19.136952	72.846822	Indian Restaurant
6	Amboli	19.1293	72.84644	Domino's Pizza	19.131000	72.848000	Pizza Place
7	Amboli	19.1293	72.84644	Garden Court	19.127188	72.837478	Indian Restaurant
8	Amboli	19.1293	72.84644	Subway	19.127860	72.844461	Sandwich Place
9	Amboli	19.1293	72.84644	Sarvodaya Veg. Restaurant	19.123760	72.850893	Indian Restaurant

Figure: Data obtained from Foursquare API after cleaning.

Methodology

This section provides details for the methodology used in the project.

Data Visualization

To understand the data obtained for Mumbai neighbourhoods, basic visualization was carried out. Figure below shows a bar plot depicting the number of neighbourhoods in each location in Mumbai.

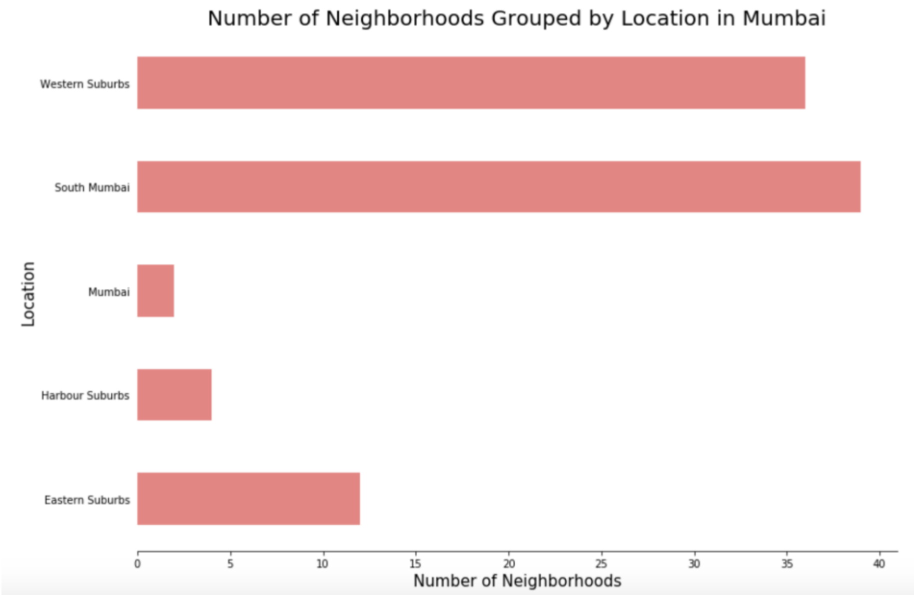

Figure: Number of neighbourhoods grouped by location.

It is evident from Figure above that South Mumbai and Western Suburbs have the greatest number of neighbourhoods. Notice how we see one of the locations as Mumbai itself? This is because the neighbourhoods contained in this location are located at the outskirts of the city and thus have been termed as just Mumbai.

Using folium, a map was plotted to show how the different neighbourhoods are spread across Mumbai. This is shown in Figure below.

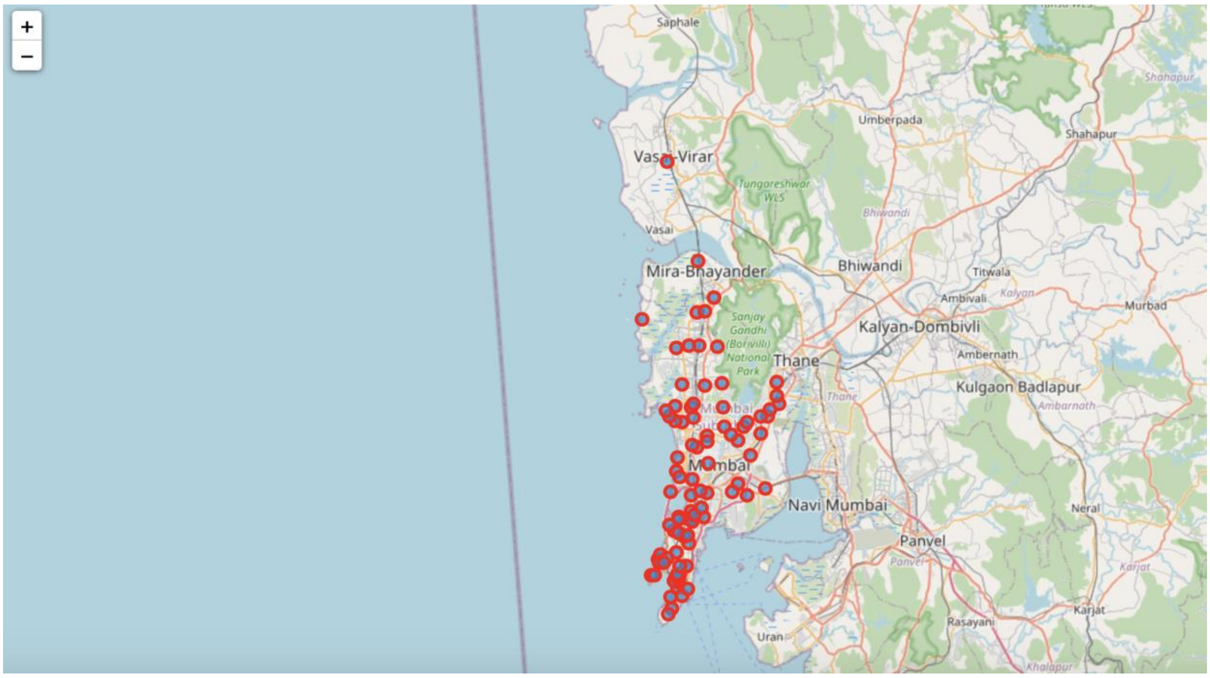

Figure: Depicting the neighbourhood spread across Mumbai.

Feature Extraction

Feature extraction was carried out to obtain features from the Foursquare API data which was used for building the unsupervised learning model. In order to achieve this, the "Venue Category" column had to be converted to some form of numeric value to be used for building the model. This was achieved by the One-hot Encoding method which takes all the unique categories and creates a column for each category. Then, if a neighbourhood venue belongs to that category, it would get a value of 1 for that row in that specific category column and if a neighbourhood venue does not belong to the category, the value would be 0. This process was repeated for all venues in all neighbourhoods and the result was a sparse matrix containing the neighbourhood's name and all unique category columns with either 1 or 0 based on whether the neighbourhood venue belonged to that category or not. This dataframe was then grouped by the neighbourhood name and the average value was taken for all categories. The result is shown in Figure below which shows only the top 10 rows.

	Neighborhood	ATM	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	...	Trail	Train	Train Station	Vegetarian / Vegan Restaurant	Whisky Bar	Wine Bar	Wine Shop	Women's Store	Yoga Studio	Zoo
0	Amboli	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
1	Chakala, Andheri	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.047619	0.0	0.0	0.000	0.000000	0.0	0.0
2	D.N. Nagar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.043478	0.0	0.0	0.000	0.021739	0.0	0.0
3	Four Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.030303	0.0	0.0	0.000	0.015152	0.0	0.0
4	Lokhandwala	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.010753	0.0	0.0	0.000	0.010753	0.0	0.0
5	Marol	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
6	Sahar	0.0	0.0	0.033333	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
7	Seven Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.014925	...	0.0	0.0	0.0	0.029851	0.0	0.0	0.000	0.000000	0.0	0.0
8	Versova	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.025000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.025	0.000000	0.0	0.0
9	Mira Road	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.066667	0.0	0.0

10 rows × 221 columns

Figure: One-hot Encoding resulting dataframe.

Notice that most of the values are 0 since there were many unique categories and not all neighbourhoods had venues belonging to each category. This data was used for the unsupervised learning model with the neighbourhood's name dropped. The unsupervised learning model is explained in the next section.

A dataframe was also created which contained the top 10 most common venues of all neighbourhoods. Though this is not a part of Feature Extraction, it is important to provide a glimpse into what this dataframe looks like as it will be used later to combine the results from the unsupervised learning model. The top 10 rows of this dataframe are shown in Figure below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Indian Restaurant	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure: Top 10 most common venues for neighbourhoods.

Unsupervised Learning

K-means unsupervised learning technique was used to cluster the neighbourhoods based on the category of venues near the neighbourhoods. One important aspect of the k-means model is to determine the number of clusters to use in model development. This was determined by the Silhouette score which was calculated for a range of clusters from 2 to 15. The resulting number of clusters and their respective Silhouette scores are shown in Figure 10.

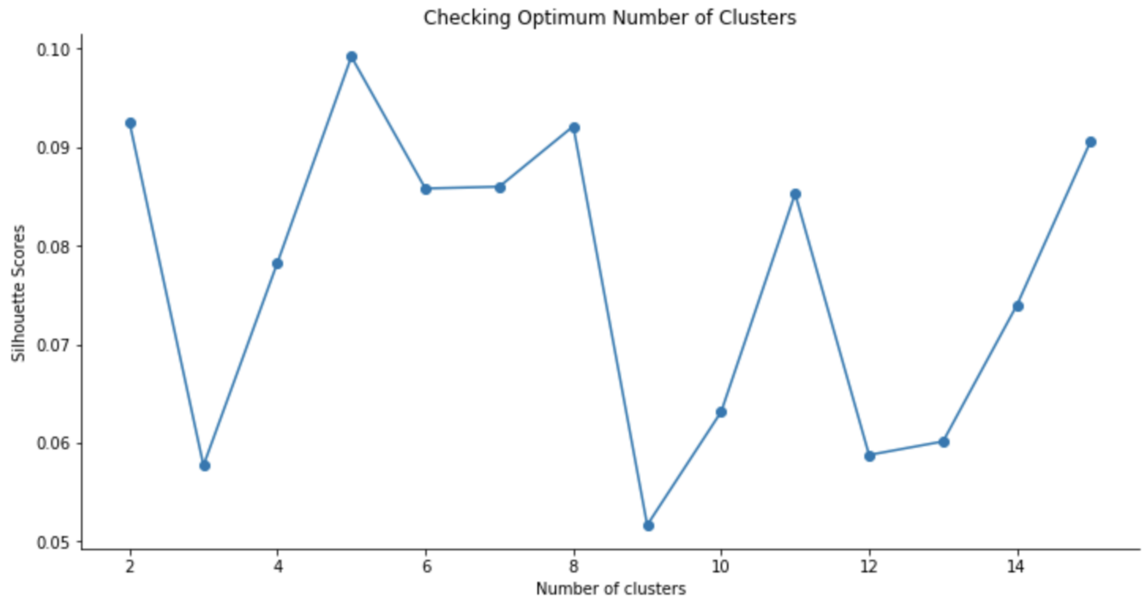

Figure: Silhouette scores for different number of clusters.

It is evident that the Silhouette scores are not very high even as the number of clusters increases. This means that the inter-cluster distance is not very high over the range of k-values. Despite this, the data will be clustered to the best possible extent. For this, 5 clusters will be used for the k-means clustering model since it provides the highest silhouette score as seen in Figure above.

Results

The clustering model then clusters the neighbourhoods in Mumbai and provides a label for each neighbourhood which is representative of the cluster it belongs to. The cluster labels were then added to the dataframe in Figure above along with the Location, Latitude, and Longitude columns to provide a complete summary of the clustering. The top 10 rows are shown in Figure below.

	Neighborhood	Location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Western Suburbs	19.1293	72.8464	1	Indian Restaurant	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	1	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	0	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Western Suburbs	19.1263	72.8243	0	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Western Suburbs	19.1432	72.8249	0	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Western Suburbs	19.1192	72.8827	1	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Western Suburbs	19.1027	72.8626	0	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Western Suburbs	19.1315	72.817	0	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Western Suburbs	19.1377	72.8135	0	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Western Suburbs	19.2657	72.8711	1	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure: Clustering neighbourhoods in Mumbai.

Furthermore, neighbourhoods in each individual cluster can be extracted using cluster labels and thus the details of specific clusters can be seen. This is done below for all clusters with only 10 rows for clusters that contain a high number of neighbourhoods.

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
32	Nalasopara	Western Suburbs	Multiplex	Bakery	Fast Food Restaurant	Ice Cream Shop	Pizza Place	Diner	Department Store	Music Venue	Music Store	Multicuisine Indian Restaurant
59	Cotton Green	South Mumbai	Plaza	Pizza Place	Fast Food Restaurant	Train Station	Multiplex	Music Venue	Modern European Restaurant	Molecular Gastronomy Restaurant	Monument / Landmark	Motorcycle Shop

Figure: Cluster 1.

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Western Suburbs	Indian Restaurant	Pizza Place	Coffee Shop	Sandwich Place	Asian Restaurant	Bar	Bowling Alley	Snack Place	Chinese Restaurant	Metro Station
1	Chakala	Western Suburbs	Hotel	Indian Restaurant	Café	Multiplex	Hotel Bar	Pizza Place	Vegetarian / Vegan Restaurant	Asian Restaurant	Fast Food Restaurant	Restaurant
5	Marol	Western Suburbs	Indian Restaurant	Hotel	Chinese Restaurant	Ice Cream Shop	Coffee Shop	Snack Place	Boat or Ferry	Fast Food Restaurant	Farmers Market	Lounge
9	Mira Road	Western Suburbs	Indian Restaurant	Convenience Store	Bar	Movie Theater	Mexican Restaurant	Multicuisine Indian Restaurant	Fast Food Restaurant	General College & University	Market	Gift Shop
11	Uttan	Western Suburbs	Beach	Convenience Store	Indian Restaurant	Whisky Bar	Restaurant	Accessories Store	Mughlai Restaurant	Music Store	Multiplex	Multicuisine Indian Restaurant
15	I.C. Colony	Western Suburbs	Indian Restaurant	Bakery	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Bar	Garden Center	Dessert Shop	Paper / Office Supplies Store	Soccer Field
20	Jogeshwari West	Western Suburbs	Indian Restaurant	Snack Place	Asian Restaurant	Ice Cream Shop	Chinese Restaurant	Bank	Hotel	Smoke Shop	Café	Mughlai Restaurant
21	Juhu	Western Suburbs	Indian Restaurant	Movie Theater	Coffee Shop	Fast Food Restaurant	Café	Vegetarian / Vegan Restaurant	Lounge	Department Store	Clothing Store	Restaurant

Figure: Cluster 2.

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
36	Bhandup	Eastern Suburbs	Train Station	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Zoo	Donut Shop	Flea Market	Fish Market	Fish & Chips Shop	Field
40	Kanjurmarg	Eastern Suburbs	Train Station	Gift Shop	Gym	Multiplex	Asian Restaurant	Chinese Restaurant	Cupcake Shop	Hotel	Field	Fast Food Restaurant
90	Dava Bazaar	South Mumbai	Train Station	Indian Restaurant	Cupcake Shop	Hotel	Fish Market	Café	Fast Food Restaurant	Coffee Shop	Asian Restaurant	Smoke Shop

Figure: Cluster 3.

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
92	Thane	Mumbai	Playground	Performing Arts Venue	Pizza Place	Music Venue	Mobile Phone Shop	Modern European Restaurant	Molecular Gastronomy Restaurant	Monument / Landmark	Motorcycle Shop	Movie Theater

Figure: Cluster 4.

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
50	Mankhurd	Harbour Suburbs	Sports Bar	Coffee Shop	Train Station	Bus Station	Accessories Store	Music Venue	Modern European Restaurant	Molecular Gastronomy Restaurant	Monument / Landmark	Motorcycle Shop
80	Dagdi Chawl	South Mumbai	Coffee Shop	Harbor / Marina	Bus Station	Train Station	Bakery	Flower Shop	Motorcycle Shop	Movie Theater	Music Venue	Moving Target

Figure: Cluster 5.

Based on the clusters shown above, the neighbourhoods can once again be plotted on a map of Mumbai, however, this time with different colour markers to distinguish between different clusters. This is shown in Figure below.

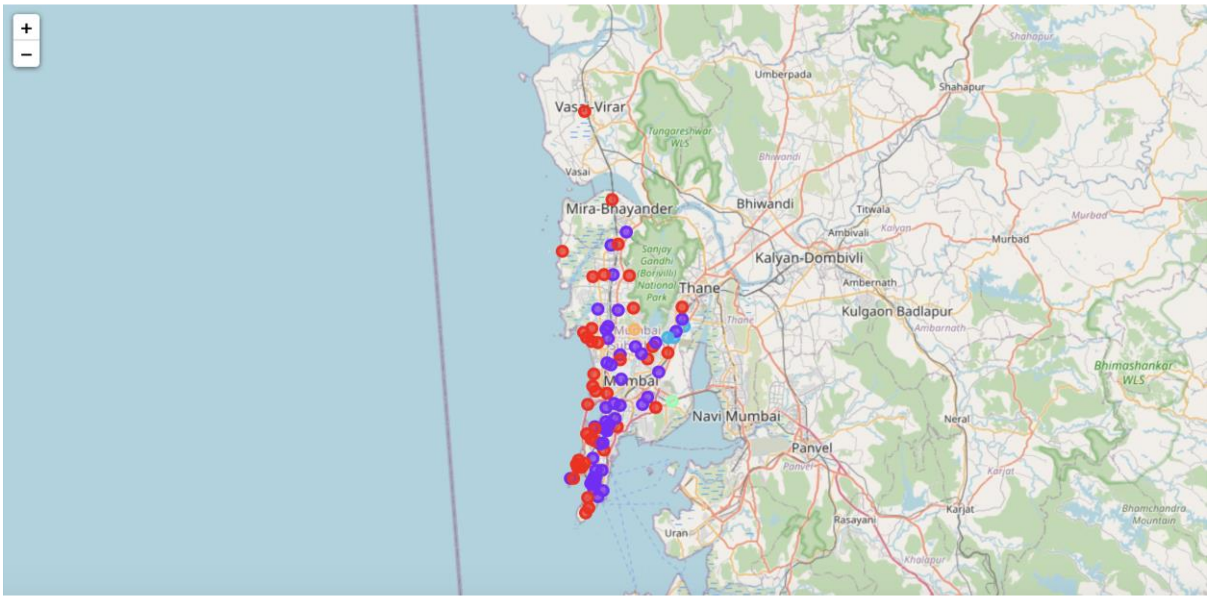

Figure: Visualizing the clustering of neighbourhoods in Mumbai.

Discussion

By analysing the five clusters obtained we can see that some of the clusters are more suited for restaurants and hotels, whereas other clusters are less suited. Neighbourhoods in clusters 3, 4, and 5 contain a small percentage of restaurants, hotels, cafe and pubs in their top 10 common venues. These clusters contain a higher degree of other venues like train station, bus station, fish market, gym, performing arts venue and smoke shop, to name a few. Thus, they are not well suited for opening a new restaurant. On the other hand, neighbourhoods in clusters 1 and 2 contain a much higher degree of restaurants, hotels, multiplex, cafes, bars and other food joints. Thus, the neighbourhoods in these clusters would be well suited for opening a new restaurant.

Comparing clusters 1 and 2, neighbourhoods in cluster 1 seem to be more suited for starting a restaurant since they contain a larger percentage of food joints in the top 10 most common venues than cluster 2. The neighbourhoods in cluster 1 contain a variety of food joints like restaurants, tea rooms, bakery, cafe, steakhouse and pubs and also contain very diverse cuisines like Japanese, Indian, Chinese, Italian and seafood restaurants. Most neighbourhoods in cluster 2 seem to have Indian Restaurant as their top most common venue; however, on careful analysis we can see that neighbourhoods in cluster 2 also contain other venues like soccer field, flea market, smoke shop, gym, train station, dance studio, music store, cosmetics shop and so on. Thus, it is recommended that the new restaurant can be opened in the neighbourhoods belonging to cluster 1. This neighbourhood can be further plotted on a map as shown below in Figure below.

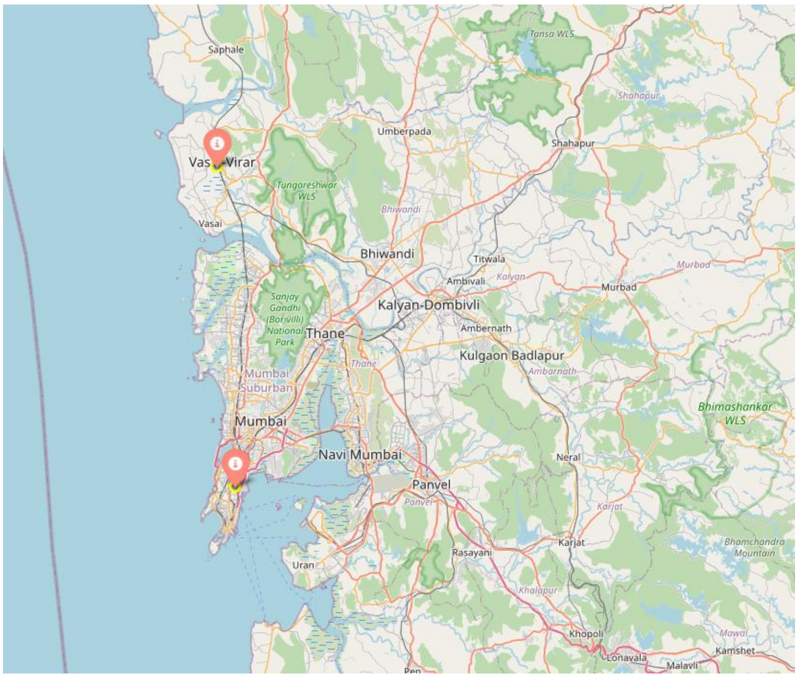

Figure: Neighbourhoods most suited for starting a new restaurant.

Conclusion

In this project, the neighbourhoods in Mumbai, India have been successfully analysed for determining which would be the best neighbourhoods for opening a new restaurant. Based on the analysis carried out, neighbourhoods in cluster 1 are recommended as locations for the new restaurant. This has also been plotted in the map in Figure below. The stakeholders and investors can further tune this by considering various other factors like transport, legal requirements, and costs associated. These were out of the scope for this project and thus were not considered.

Final Comments

To view the code for this project, kindly refer to the notebook on the GitHub repository:https://github.com/Tanveen11/Coursera_Capstone/blob/87bc1e019a1b2b840a7045ecf749ef4722292f83/Final%20Notebook.ipynb